The Leiden Ranking 2011/2012: Data Collection, Indicators, and Interpretation¹


waltmanlr@cwts.leidenuniv.nl
Centre for Science and Technology Studies, Leiden University (The Netherlands)

Abstract
The Leiden Ranking 2011/2012 is a ranking of universities based on bibliometric indicators of publication output, citation impact, and scientific collaboration. The ranking includes 500 major universities worldwide. This paper provides an extensive discussion of the Leiden Ranking 2011/2012. A detailed description is offered of the indicators used in the ranking, and various innovations in the Leiden Ranking 2011/2012 are presented. These innovations include (1) an indicator based on counting a university’s highly cited publications, (2) indicators based on fractional rather than full counting of collaborative publications, (3) the possibility of excluding non-English language publications, and (4) the use of stability intervals. The paper also comments on the interpretation of the Leiden Ranking 2011/2012. A number of limitations of the ranking are pointed out as well.

Introduction
The Leiden Ranking is a global university ranking based exclusively on bibliometric data. In this paper, we introduce the 2011/2012 edition of the Leiden Ranking. The paper provides a discussion of the data collection methodology, the indicators used in the ranking, and the interpretation of the ranking. The Leiden Ranking 2011/2012 is available on the website www.leidenranking.com.

University rankings have quickly gained popularity, especially since the launch of Academic Ranking of World Universities, also known as the Shanghai Ranking, in 2003, and these rankings nowadays play a significant role in university decision making. The increased use of university rankings has not been hampered by the methodological problems that were already identified in an early stage (e.g., Van Raan, 2005). There are now many rankings in which universities are compared on one or more dimensions of their performance. Many of these rankings have a national or regional focus, or they consider only specific scientific disciplines. There is a small group of global university rankings (Rauhvargers, 2011). The Leiden Ranking belongs to this group of rankings.

Global university rankings are used for a variety of purposes by different user groups. Three ways of using university rankings seem to be dominant. First, governments, funding agencies, and the media use university rankings as a source of strategic information on the global competition among universities. Second, university managers use university rankings as a marketing and decision support tool. And third, students and their parents use university rankings as a selection instrument.

¹ We would like to thank Henk Moed for his contribution to earlier editions of the Leiden Ranking.
An important methodological problem of the most commonly used global university rankings is their combination of multiple dimensions of university performance in a single aggregate indicator. These dimensions, which often relate to very different aspects of university performance (e.g., scientific performance and teaching performance), are combined in a quite arbitrary fashion. This prevents a clear interpretation of the aggregate indicator. A second related problem has to do with the fact that different universities may have quite different missions. Two universities that each have an excellent performance on the dimension that is most relevant to their mission may end up at very different positions in a ranking if the different dimensions are weighted differently in the aggregate indicator. These methodological problems can partly be solved by providing separate scores on the various dimensions and refraining from aggregating these scores in a single number. A third problem is more practical. Some rankings rely heavily on data supplied by the universities themselves, for instance data on staff numbers or student/staff ratios. This dependence on the universities makes these rankings vulnerable to manipulation. Also, because of the lack of internationally standardized definitions, it is often unclear to what extent data obtained from universities can be used to make valid comparisons across universities or countries.

A solution to these fundamental methodological problems is to restrict a ranking to a single dimension of university performance that can be measured in an accurate and reliable way. This is the solution that the Leiden Ranking offers. The Leiden Ranking does not attempt to measure all relevant dimensions of university performance. Instead, the ranking restricts itself to the dimension of scientific performance. Other dimensions of university performance, in particular the dimension of teaching performance, are not considered. The Leiden Ranking includes 500 major universities worldwide and is based on bibliometric data from the Web of Science database. No data is employed that has been supplied by the universities themselves. A sophisticated procedure for assigning publications to universities is used to further improve the quality of the bibliometric data.

The first edition of the Leiden Ranking was produced in 2007. In this paper, we discuss the 2011/2012 edition of the Leiden Ranking. This edition was published in December 2011 on www.leidenranking.com. Compared with earlier editions of the Leiden Ranking, the 2011/2012 edition offers a number of innovations. These innovations address some of the shortcomings of earlier editions of the ranking and also of other university rankings. Below, we summarize the most important innovations:

- The $PP_{top\,10\%}$ indicator has been added to the Leiden Ranking. Compared with other citation impact indicators, an important advantage of the $PP_{top\,10\%}$ indicator is its insensitivity to extremely highly cited publications.

- The fractional counting method has been added to the Leiden Ranking. We argue that, compared with the more traditional full counting method, the fractional counting method leads to more accurate comparisons between universities.

- The possibility of excluding non-English language publications has been added to the Leiden Ranking. These publications may disadvantage universities from, for instance, France and Germany.

- Stability intervals have been added to the Leiden Ranking. A stability interval provides insight into the sensitivity of an indicator to changes in the underlying set of publications.

The above innovations are discussed in more detail later on in this paper.

The rest of the paper is organized as follows. We first briefly describe the data collection
methodology of the Leiden Ranking. We then extensively discuss the indicators that are used in the ranking, paying special attention to the innovations that have been made in the 2011/2012 edition of the ranking. Next, we comment on the interpretation of the Leiden Ranking. Finally, we conclude the paper and discuss our future plans for the Leiden Ranking. We note that a more extensive version of this paper is available as a working paper (Waltman & al., 2012). In this more extensive version, we also compare the Leiden Ranking with other global university rankings and we describe the data collection methodology of the Leiden Ranking in a more comprehensive way. In addition, we offer a more extensive empirical analysis.

Data collection
As already mentioned, the Leiden Ranking limits itself to universities only. Other types of research institutions are not considered. Data on the publications of universities was collected from Thomson Reuters’ Web of Science (WoS) database. We only considered publications of the document types article, letter, and review that were published between 2005 and 2009. Also, we only included publications from the sciences and the social sciences. Publications with an arts and humanities classification in WoS were excluded. Our focus was on universities with at least 500 publications in each of the five years. Changes in the organizational structure of universities were taken into consideration up to 2009. Mergers, split-ups, and other changes that took place after 2009 may not have been taken into account.

Due to space limitations, it is not possible to discuss in any detail the procedure that we used for assigning publications to universities. We refer to our working paper (Waltman & al., 2012) for an extensive discussion of this procedure. In the end, the 500 universities for which the largest WoS publication output was obtained were included in the Leiden Ranking.

Indicators
The Leiden Ranking provides three types of indicators: Indicators of publication output, indicators of citation impact, and indicators of scientific collaboration. Publication output is measured using the number of publications (P) indicator. This indicator is calculated by counting the total number of publications of a university. Publications that have the document type letter in WoS do not count as a full publication but count as one fourth of a publication. The indicators used to measure impact and collaboration are discussed in the first two subsections below.

In the next three subsections, we consider three special feature of the Leiden Ranking: (1) The two counting methods supported by the ranking (i.e., full counting and fractional counting), (2) the possibility of excluding non-English language publications from the calculation of the indicators, and (3) the possibility of complementing indicators with so-called stability intervals.

Unless stated otherwise, the empirical results reported in this section were obtained using the fractional counting method based on English-language publications only. We also note that all results of the Leiden Ranking are available in an Excel file that can be downloaded from www.leidenranking.com.

Impact indicators

---

2 Counting letters as one fourth of an ordinary publication (i.e., an article or a review) of course involves some arbitrariness. We have chosen to use a weight of 0.25 for letters because in WoS a letter on average receives roughly one fourth of the citations of an ordinary publication.
The Leiden Ranking includes three indicators of the citation impact of the work of a university:

- **Mean citation score** (MCS). The average number of citations of the publications of a university.
- **Mean normalized citation score** (MNCS). The average number of citations of the publications of a university, normalized for differences between scientific fields (i.e., WoS subject categories), differences between publication years, and differences between document types (i.e., article, letter, and review). An MNCS value of one can be interpreted as the world average (or more properly, the average of all WoS publications). Consequently, if a university has an MNCS value of two, this for instance means that the publications of the university have been cited twice above world average. We refer to Waltman, Van Eck, Van Leeuwen, Visser and Van Raan (2011) for a more detailed discussion of the MNCS indicator.
- **Proportion top 10% publications** (PP_{top 10%}). The proportion of the publications of a university that, compared with other similar publications, belong to the top 10% most frequently cited. Publications are considered similar if they were published in the same field and the same publication year and if they have the same document type. We note that an indicator similar to our PP_{top 10%} indicator was recently introduced in the Scimago Institutions Rankings (Bornmann, De Moya-Anegón & Leydesdorff, 2012).

In the calculation of the above indicators, citations are counted until the end of 2010. Author self citations are excluded from all calculations.³ Publications of the document type letter are weighted as one fourth of a full publication.

Figure 1 shows a scatter plot of the relation between the MNCS indicator and the PP_{top 10%} indicator for the 500 Leiden Ranking universities. There is a strong, more or less linear relation between the two indicators. However, there is one university for which the indicators deviate strongly from this relation. This is University of Göttingen. This university is ranked 2nd based on the MNCS indicator, while it is ranked 238th based on the PP_{top 10%} indicator. The MNCS indicator for University of Göttingen turns out to have been strongly influenced by a single extremely highly cited publication. This publication (Sheldrick, 2008) was published in January 2008 and had been cited over 16,000 times by the end of 2010. Without this single publication, the MNCS indicator for University of Göttingen would have been equal to 1.09 instead of 2.04, and University of Göttingen would have been ranked 219th instead of 2nd. Unlike the MNCS indicator, the PP_{top 10%} indicator is hardly influenced by a single very highly cited publication. This is because the PP_{top 10%} indicator only takes into account whether a publication belongs to the top 10% of its field or not. The indicator is insensitive to the exact number of citations of a publication. This is an important difference with the MNCS indicator, and this difference explains why in the case of University of Göttingen the MNCS indicator and the PP_{top 10%} indicator yield very different results. In our view, the sensitivity of the MNCS indicator to a single very highly cited publication is an undesirable property. We therefore regard the PP_{top 10%} indicator as the most important impact indicator in the Leiden Ranking.

---

³ A citation is regarded as an author self citation if the citing and the cited publication have at least one author name (i.e., last name and initials) in common. We tested the effect of excluding self citations on the MNCS and PP_{top 10%} indicators. For most universities, the effect turns out to be negligible. There is a small set of universities for which the effect is more substantial. These are mainly universities from continental Europe, especially from Germany. For these universities, excluding self citations considerably decreases the MNCS and PP_{top 10%} indicators.
Collaboration indicators

The Leiden Ranking includes four indicators of the degree to which a university is involved in scientific collaborations with other organizations:

- **Proportion collaborative publications (PP\textsubscript{collab})**. The proportion of the publications of a university that have been co-authored with one or more other organizations.
- **Proportion international collaborative publications (PP\textsubscript{int collab})**. The proportion of the publications of a university that have been co-authored by two or more countries.
- **Mean geographical collaboration distance (MGCD)**. The average geographical collaboration distance of the publications of a university. The geographical collaboration distance of a publication is defined as the largest geographical distance between two addresses mentioned in the publication’s address list. If a publication’s address list contains only one address, the geographical collaboration distance of the publication equals zero. We refer to Tijssen, Waltman and Van Eck (2011) and Waltman, Tijssen and Van Eck (2011) for a more detailed discussion of the MGCD indicator, including a discussion of the geocoding procedure that was used to identify the geographical coordinates of the addresses mentioned in publications’ address lists.\(^4\)
- **Proportion long distance collaborative publications (PP\textsubscript{$>$1000 km})**. The proportion of the publications of a university that have a geographical collaboration distance of more than 1000 km.

Like in the impact indicators discussed above, publications of the document type *letter* are weighted as one fourth of a full publication in the above indicators.

---

\(^4\) We note that there are some small inaccuracies in the calculation of the MGCD indicator. This is because for some of the addresses mentioned in publications’ address lists we do not have the geographical coordinates. In the case of the Leiden Ranking universities, about 2.3% of the publications have at least one address without geographical coordinates. Addresses without geographical coordinates are ignored in the calculation of the MGCD indicator.
**Counting method**

The Leiden Ranking supports two counting methods: Full counting and fractional counting. In the calculation of the indicators, the full counting method gives equal weight to all publications of a university (except for publications of the document type letter). The fractional counting method gives less weight to collaborative publications than to non-collaborative ones. For instance, if the address list of a publication contains five addresses and two of these addresses belong to a particular university, then the publication has a weight of 0.4 in the calculation of the indicators for this university. Using the fractional counting method, a publication is fully assigned to a university only if all addresses mentioned in the publication’s address list belong to the university.

For the purpose of making comparisons between universities, we consider the fractional counting method preferable over the full counting method. This is based on the following argument. If for each publication in WoS we calculate the MNCS indicator, the average of all these publication-level MNCS values will be equal to one. We want a similar property to hold at the level of organizations. If each publication in WoS belongs to one or more organizations and if for each organization we calculate the MNCS indicator, we want the average (weighted by publication output) of all these organization-level MNCS values to be equal to one. If this property holds, the value one can serve as a benchmark not only at the level of publications but also at the level of organizations. This would for instance mean that an organization with an MNCS indicator of two can be said to perform twice above average in comparison with other organizations. Using the full counting method, however, the above property does not hold. This is because publications belonging to multiple organizations are fully counted multiple times, once for each organization to which they belong. This double counting of publications causes the average of the organization-level MNCS values to deviate from one. Using the fractional counting method, on the other hand, it can be shown that the above property does hold. Therefore, if the fractional counting method is used, the value one can serve as a proper benchmark at the organization level. This is our main argument for preferring the fractional counting method over the full counting method.5 We note that the argument is not restricted to the MNCS indicator. The argument also applies to other indicators, such as the PP top 10% indicator.

In practice, the full counting method causes the average of the organization-level MNCS values to be greater than one. Similarly, it causes the average of the organization-level PP top 10% values to be greater than 10%. This is due to a combination of two mechanisms. First, collaborative publications are counted multiple times in the full counting method, and second, collaborative publications tend to be cited more frequently than non-collaborative publications. The combination of these two mechanisms is responsible for the effect that at the level of organizations MNCS and PP top 10% values on average are greater than, respectively, one and 10%. Importantly, there are substantial differences between scientific fields in the strength of this effect. For instance, the effect is very strong in clinical medicine and quite weak in chemistry, engineering, and mathematics.6 Because of

---

5 A similar argument in favor of the fractional counting method is given in a recent paper by Aksnes, Schneider and Gunnarsson (2012), in which the full counting method and the fractional counting method are compared at the level of countries.

6 This statement is based on the following analysis. For different scientific fields, we calculated the average number of citations per publication. Both a weighted and an unweighted average were calculated. In the case of the weighted average, each publication was weighted by the number of addresses in the address list. The ratio of the weighted and the unweighted average provides an indication of the size of the ‘full counting bonus’. A ratio of approximately 1.35 was obtained for clinical medicine. Ratios below 1.10 were obtained for chemistry, engineering, and mathematics.
these differences between fields, the full counting method may be considered biased in favor of some organizations over others. Organizations active mainly in clinical medicine research for instance have an advantage over organizations focusing on engineering research.

Figure 2 shows a scatter plot of the relation between the PP\textsubscript{top 10\%} indicator calculated using the full counting method and the PP\textsubscript{top 10\%} indicator calculated using the fractional counting method. For almost all universities, the PP\textsubscript{top 10\%} indicator calculated using the full counting method has a higher value than the PP\textsubscript{top 10\%} indicator calculated using the fractional counting method. This is a consequence of the ‘full counting bonus’ discussed above. The overall correlation between the full counting method and the fractional counting method turns out to be high ($r = 0.97$), but as can be seen in Figure 2, at the level of individual universities the difference between the two counting methods can be quite significant. Tables 1 and 2 list the five universities that, based on the PP\textsubscript{top 10\%} indicator, benefit most from either the full or the fractional counting method. In line with the above discussion, the universities benefiting from the full counting method almost all have a clear medical profile. (University of Nantes is an exception.) The other way around, the universities benefiting from the fractional counting method all have a strong focus on engineering research and on the natural sciences. Most of these universities are located in Asia.

**Table 1.** The five universities that, based on the PP\textsubscript{top 10\%} indicator, benefit most from the full counting method.

<table>
<thead>
<tr>
<th>University</th>
<th>Country</th>
<th>PP\textsubscript{top 10%} indicator</th>
<th>Full counting</th>
<th>Fractional counting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lille 2 University of Health and Law</td>
<td>France</td>
<td>15.6%</td>
<td>9.9%</td>
<td></td>
</tr>
<tr>
<td>Wake Forest University</td>
<td>United States</td>
<td>16.8%</td>
<td>12.0%</td>
<td></td>
</tr>
<tr>
<td>Hannover Medical School</td>
<td>Germany</td>
<td>14.1%</td>
<td>10.0%</td>
<td></td>
</tr>
<tr>
<td>University of Nantes</td>
<td>France</td>
<td>13.5%</td>
<td>9.4%</td>
<td></td>
</tr>
<tr>
<td>University of Alabama at Birmingham</td>
<td>United States</td>
<td>14.9%</td>
<td>11.0%</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. The five universities that, based on the PP\textsubscript{top 10\%} indicator, benefit most from the fractional counting method.

<table>
<thead>
<tr>
<th>University</th>
<th>Country</th>
<th>PP\textsubscript{top 10%} indicator</th>
<th>Full counting</th>
<th>Fractional counting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nankai University</td>
<td>China</td>
<td>12.7%</td>
<td></td>
<td>13.4%</td>
</tr>
<tr>
<td>Rice University</td>
<td>United States</td>
<td>21.7%</td>
<td></td>
<td>22.2%</td>
</tr>
<tr>
<td>Pohang University of Science and Technology</td>
<td>South Korea</td>
<td>13.7%</td>
<td></td>
<td>14.1%</td>
</tr>
<tr>
<td>Indian Institute of Technology Kharagpur</td>
<td>India</td>
<td>8.7%</td>
<td></td>
<td>9.0%</td>
</tr>
<tr>
<td>National Chung Hsing University</td>
<td>Taiwan</td>
<td>9.2%</td>
<td></td>
<td>9.4%</td>
</tr>
</tbody>
</table>

Non-English language publications

About 2.1\% of the publications of the Leiden Ranking universities have not been written in English. Of these non-English language publications, most have been written in German (31\%), Chinese (17\%), French (17\%), Spanish (13\%), or Portuguese (10\%). Comparing the impact of non-English language publications with the impact of publications written in English may not be considered fair (Van Raan, Van Leeuwen & Visser, 2011a, 2011b). Non-English language publications can be read only by a small part of the scientific community, and therefore these publications cannot be expected to receive similar numbers of citations as publications written in English. To deal with this issue, the Leiden Ranking offers the possibility of excluding non-English language publications from the calculation of the indicators.

Table 3. The five universities that, based on the PP\textsubscript{top 10\%} indicator, benefit most from excluding non-English language publications.

<table>
<thead>
<tr>
<th>University</th>
<th>Country</th>
<th>PP\textsubscript{top 10%} indicator</th>
<th>All publications</th>
<th>English pub. only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lille 2 University of Health and Law</td>
<td>France</td>
<td>6.9%</td>
<td>9.9%</td>
<td></td>
</tr>
<tr>
<td>Université Bordeaux Segalen</td>
<td>France</td>
<td>10.2%</td>
<td>12.5%</td>
<td></td>
</tr>
<tr>
<td>Montpellier I University</td>
<td>France</td>
<td>8.7%</td>
<td>10.7%</td>
<td></td>
</tr>
<tr>
<td>Paris Descartes University</td>
<td>France</td>
<td>9.9%</td>
<td>11.9%</td>
<td></td>
</tr>
<tr>
<td>Université de la Méditerranée - Aix-Marseille II</td>
<td>France</td>
<td>8.4%</td>
<td>10.0%</td>
<td></td>
</tr>
</tbody>
</table>

The overall correlation between the PP\textsubscript{top 10\%} indicator based on all publications and the PP\textsubscript{top 10\%} indicator based on English-language publications only is very high ($r = 0.99$), and for most universities including or excluding non-English language publications makes hardly any difference. Nevertheless, there are a number of universities that benefit quite significantly from excluding non-English language publications. These are mostly French and German universities, but also some from China and other countries. Table 3 lists the five universities that, based on the PP\textsubscript{top 10\%} indicator, benefit most from excluding non-English language publications. We refer to Van Raan & al. (2011a) for additional empirical results on the effect of excluding non-English language publications.

Stability intervals

The stability of an indicator relates to the sensitivity of the indicator to changes in the underlying set of publications. An indicator has a low stability if it is highly sensitive to changes in the set of publications based on which it is calculated. An indicator has a high stability if it is relatively insensitive to such changes. For instance, if a university has one or a few very highly cited publications and a large number of lowly cited publications, the MNCS indicator for this university will be relatively unstable. This is because the value of the MNCS indicator depends
strongly on whether the university’s highly cited publications are included in the calculation of the indicator or not. A university whose publications all have similar citation scores will have a very stable MNCS indicator. In general, the larger the number of publications of a university, the more stable the indicators calculated for the university.

To provide some insight into the stability of indicators, the Leiden Ranking uses so-called stability intervals. Stability intervals are similar to confidence intervals, but they have a somewhat different interpretation. A stability interval indicates a range of values of an indicator that are likely to be observed when the underlying set of publications changes. For instance, the MNCS indicator may be equal to 1.50 for a particular university, with a stability interval from 1.40 to 1.65. This means that the true value of the MNCS indicator equals 1.50 for this university, but that changes in the set of publications of the university may relatively easily lead to MNCS values in the range from 1.40 to 1.65. The larger the stability interval of an indicator, the lower the stability of the indicator.

The stability intervals used in the Leiden Ranking are constructed as follows. Consider a university with \( n \) publications, and suppose we want to construct a stability interval for the MNCS indicator of this university. We then randomly draw 1,000 samples from the set of publications of the university. Each sample is drawn with replacement, which means that a publication may occur multiple times in the same sample. The size of each sample is \( n \), which is equal to the number of publications of the university. For each sample, we calculate the value of the MNCS indicator. This yields a distribution of 1,000 sample MNCS values. We use this distribution to determine a stability interval for the MNCS indicator of the university. The Leiden Ranking uses 95% stability intervals. To obtain a 95% stability interval, we take the 2.5th and the 97.5th percentile of the distribution of sample MNCS values. These percentiles serve as the lower and the upper bound of the stability interval.

In the statistical literature, the above procedure for constructing stability intervals is known as bootstrapping (Efron & Tibshirani, 1993). Stability intervals are also discussed in a recent paper by Colliander and Ahlgren (2011). However, Colliander and Ahlgren use a somewhat different procedure for constructing stability intervals than we do.7

**Interpretation of the Leiden Ranking**

University rankings aim to capture a complex reality in a small set of numbers. By necessity, this imposes serious limitations on these rankings. Below, we summarize a number of important limitations that should be taken into account in the interpretation of the Leiden Ranking:

1. The Leiden Ranking is based exclusively on output variables of the process of scientific research (i.e., publications, citations, and co-authorships). Input variables, such as the number of research staff of a university or the amount of money a university has available for research, are not taken into account. Ideally, scientific performance should be measured based on both input and output variables (see also Calero-Medina, López-Illescas, Visser & Moed, 2008). However, accurate internationally standardized data on input variables is not available, and this is why the Leiden Ranking uses output variables only.

---

7 Leydesdorff and Bornmann (in press) suggest the use of a statistical test to determine whether differences between universities in the Leiden Ranking are statistically significant. This can be seen as an alternative to our stability intervals. However, given the various problems associated with statistical tests (Schneider, 2011), we prefer the use of our stability intervals.
2. Indicators like those used in the Leiden Ranking (but also in other university rankings) can be quite sensitive to all kinds of choices regarding the details of their calculation. This is well illustrated by the empirical results presented earlier in this paper, which show the effect of the choice of a counting method and the effect of the choice to include or exclude non-English language publications. The sensitivity of indicators to choices like these should be kept in mind in the interpretation of the Leiden Ranking. Many choices are somewhat hidden in the details of the calculation of an indicator and may appear to be of a rather technical nature. Nevertheless, these choices may significantly affect the results produced by an indicator. In a sense, what this means is that the results produced by an indicator are always subject to a certain degree of uncertainty. If the indicator had been calculated in a slightly different way, the results would have looked differently. It is important to emphasize that this type of uncertainty is difficult to quantify and is not reflected in the stability intervals discussed in the previous section. Stability intervals only reflect uncertainty related to changes in the set of publications underlying an indicator.

3. In the interpretation of university rankings, attention often focuses almost completely on the ranks of universities (e.g., “University X is ranked 20 positions higher than university Y” or “Country Z has five universities in the top 100”). This type of interpretation has the advantage of being easy to understand by a broad audience. However, the interpretation can also be misleading. Using indicators such as MNCS or PP_{top 10\%}, the performance of universities tends to be quite skewed. There are a small number of universities with a very high performance (e.g., MNCS above 1.8) and a large number of universities with a more average performance (e.g., MNCS between 1.0 and 1.5). This for instance means that according to the MNCS indicator the difference between the universities on ranks 1 and 10 in the Leiden Ranking is almost 0.5 while the difference between the universities on ranks 200 and 300 is less than 0.1. In other words, an increase in the rank of a university by, say, ten positions is much more significant in the top of the ranking than further down the list. A more accurate interpretation of university rankings in general and of the Leiden Ranking in particular can be obtained by looking directly at the values of the indicators rather than at the rankings implied by these values. For instance, “University X is performing 20\% better than university Y” is more accurate than “University X is ranked 20 positions higher than university Y.”

4. The Leiden Ranking assesses universities as a whole and therefore cannot be used to draw conclusions regarding the performance of individual research groups, departments, or institutes within a university. Different units within the same university may differ quite a lot in their performance, and drawing conclusions at the level of individual units based on the overall performance of a university is therefore not allowed. More detailed bibliometric analyses are needed to draw conclusions at the level of individual units within a university.

Conclusion
In this paper, we have introduced the Leiden Ranking 2011/2012. We have discussed the data collection methodology, the indicators used in the ranking, and the interpretation of the ranking.

Compared with other global university rankings, in particular the popular Shanghai and Times Higher Education rankings, the Leiden Ranking offers a number of important advantages. First, the Leiden Ranking refrains from arbitrarily combining multiple dimensions of university performance
in a single aggregate indicator. Second, the Leiden Ranking does not rely on data supplied by the universities themselves and also does not use questionable survey data. And third, the Leiden Ranking is extensively documented, making it more transparent than many other rankings.

At the same time, we also acknowledge a number of limitations of the Leiden Ranking. Depending on the purpose for which a university ranking is used, the exclusive focus of the Leiden Ranking on scientific performance can be a serious limitation. Because of this limitation, the Leiden Ranking is not very useful for prospective undergraduate students in their choice of a university. The Leiden Ranking captures the scientific performance of a university mainly by measuring the citation impact of the university’s publications. This also involves some limitations. On the one hand, citation impact is only one element of scientific performance. It does not capture elements such as the societal impact of the work of a university. On the other hand, the measurement of citation impact has various methodological difficulties, for instance because of restrictions imposed by the Web of Science database, because of limitations of the indicators that are used, and because of intrinsic difficulties associated with certain scholarly disciplines (e.g., humanities and some of the social sciences). Another shortcoming of the Leiden Ranking is the absence of a disciplinary breakdown. The Leiden Ranking offers statistics only at the level of science as a whole. Clearly, for many purposes, more fine-grained statistics are needed, for instance at the level of individual scientific fields. Such statistics are not available in the Leiden Ranking, but they can be calculated as part of performance analyses for specific universities.

We plan to further extend the Leiden Ranking in the next editions. We are considering extensions in three directions. First, the number of universities included in the Leiden Ranking may be increased, and other types of research institutions may be added to the ranking. Also, a classification of universities (e.g., ‘general university’, ‘medical university’, ‘technical university’, etc.) may be developed in order to facilitate comparisons among similar entities. Second, the statistics offered by the Leiden Ranking may be refined, for instance by reporting longitudinal trends and by providing a breakdown into a number of broad scientific disciplines. And third, the indicators used in the Leiden Ranking may be improved, and new indicators may be added. For instance, there may be room for a more sophisticated approach to the normalization of impact indicators for field differences, and an indicator of university-industry collaboration may be added to the ranking.

References


